What Is Claimed Is:

1.   A database comprising a plurality of records, said records containing phenotype information and optionally sample information for an individual, wherein the record for the individual further comprises confounding information, and the sample information for the individual comprises information relating to the location of a sample of tissue or of fluid from the individual.

2.   A database according to Claim 1, wherein the record for an individual comprises information relating to a plurality of phenotypes and the record comprises, in respect of each phenotype:-

the phenotype observed; and

information relating to actual or potential confounding indicators in respect of phenotype.

3.   A database according to Claim 1, wherein said confounding information is selected from information selected from the group consisting of medication being taken by the individual, medical history, occupational information, information relating to the hobbies of the individual, diet information, family history, normal exercise routines of the individual, age and sex.

4.   A database according to Claim 1, wherein the phenotype and confounding information is collected at the same time from the individual.

5.   A database according to Claim 1, comprising a plurality of records, each record containing genotype information, and optionally sample information for an individual, wherein:

the phenotype information for the individual comprises at least one of and optionally all of osteoporosis related phenotypes, osteoarthritis related phenotypes, immune cell subtypes (such as Tcell subsets), metabolic syndrome/syndrome X related phenotypes, and hypertension related phenotypes; and

the sample information for individual comprises information relating to the location of a sample of tissue or of fluid from the individual.

6. A database according to Claim 5, wherein the phenotype information further comprises at least one of and optionally all of thrombosis/fibrinolysis phenotypes, haemoglobinopathy related phenotypes and airways disease (asthma) phenotype.

7. A database according to Claim 6, wherein the phenotype information further comprises information relating to one or more of the phenotypes: atopy/eczema, lung function, IgE, psoriasis, acne, skin cancer and moliness of skin.

8. A database according to Claim 1 comprising a plurality of records for human individuals.

9. A database according to Claim 1 wherein the sample of tissue or of fluid is selected from the group consisting of urine, serum, skin, liver, heart, bone, hair, muscle, kidney, tooth, saliva, faeces and DNA.

10. A database according to Claim 1 wherein the sample information comprises the geographical location of the sample, the storage conditions of the sample and the storage reference number for reference label of the sample.

11.   A database according to Claim 10 wherein the sample information additionally comprises contact information enabling the individual to be contacted and retested in person.

12.   A database according to Claim 1, wherein each record further includes genotype information for the individual comprising one or more single nucleotide polymorphisms.

13.   A database according to Claim 1, comprising genotype information selected from one or more of:

(i)   actual or inferred DNA base sequence at one or more regions within the genome;

(ii)   a record of variation between a specified sequence on a chromosome of that individual compared to a reference sequence; and

(iii)   length of a particular sequence or a particular sequence variant.

14.   A method of adding information to a database according to Claim 1 comprising:

(1)   identifying an individual not yet included in the database;

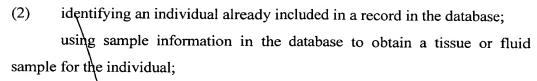determining phenotype information for the individual;

determining confounding information in respect of that phenotype information for the individual;

optionally determining genotype information for the individual;

optionally determining sample information for the individual that includes information relating to the location of the sample of tissue or of fluid from the individual; and

creating a record in the database to hold the phenotype, confounding and optionally genotype and/or sample information for the individual;

or

(2)   identifying an individual already included in a record in the database;

using sample information in the database to obtain a tissue or fluid sample for the individual;

testing the sample, thereby determining genotype or phenotype information for the individual; and

adding or confirming or amending or updating information in the record for the individual.

15.   A method of identifying a correlation between phenotype information and genotype information comprising:

selecting a phenotype characteristic;

identifying a plurality of records from the database of Claim 1 for individuals that comply with the selected phenotype characteristic; and

taking account of the confounding information, determining if presence of the selected phenotype characteristic is correlated with presence of any genotype characteristic in the genotype information for records in the database.

16.   A method of identifying a correlation between first phenotype information and second phenotype information comprising:

selecting a first phenotype characteristic;

identifying a plurality of records in the database of Claim 1 for individuals who comply with the first phenotype information;

determining if presence of the selected first phenotype is correlated with second phenotype information of records in the database.

17.   A method of identifying a correlation between genotype information and genotype information comprising:

selecting a genotype characteristic;

identifying a plurality of records in the database of claim 1 for individuals who comply with the genotype characteristic;

determining if presence of the selected genotype characteristic is correlated with another characteristic of genotype information or records in the database.

18.   A method of allocating priority to a candidate gene or locus, proposed as a drug target for treatment of a disease, the method comprising:-

calculating, from data on a database according to Claim 1, the specificity of the candidate gene or locus for the disease;

comparing (i) the association of the disease with clinical risk traits related to the disease, to (ii) the association of the disease with other clinical risk traits unrelated to the disease, but representing significant side effects; and

hence calculating a likely therapeutic index of drug candidates acting on that gene or locus.

19.   A method of analysing the relation between a genotype and a phenotype, comprising

selecting a phenotype characteristic;

identifying a plurality of records in a database according to Claim 1 complying with that characteristic;

using environmental and age-related data in the database to eliminate the effects of age and environment on variations in phenotype; and

hence calculating from the database whether and if so to what extent the phenotype is correlated with a particular genotype.

20.   A method of determining the capacity and specificity of a genetic marker to detect and quantify normal variations in healthy and affected populations for a selected risk trait, comprising:-

assaying samples in a database according to Claim 1 for the marker levels, in both healthy and affected subjects; and

quantifying the association of the clinical trait with the marker level and other selected phenotypes, in unaffected and affected subjects.

21.   A method of devising dose regimes and/or dose forms and/or drug delivery systems for a given drug in a clinical trial, comprising:-

selecting a proposed clinical population for the trial;

using data on a database according to Claim 1 to stratify the clinical population by high associations of metabolism or absorption of the drug both with genotype and/or with associated biochemical and cell biology phenotypes; and

hence allowing definition of the best dose regimes and dose forms/drug delivery systems;

so as to predict and/or allow for absorption and/or metabolism of the drug by patients in the clinical population.

22.   A method of predicting response to a proposed drug therapy, comprising:-

using a database according to Claim 1 to select a clinical population by constructing haplotypic profiles, with strong associations with defined clinical traits and biochemical phenotypes;

using the database to eliminate the effects of age and environment in the clinical population;

hence providing criteria to predict response to the drug and variation in response to the drug, and optionally to define a sub-group of the clinical population or of the general population most susceptible to the drug being studied.

23.   A method of correlating genotype and phenotype information with account taken of potential or actual confounding information, comprising use of the database of Claim 1.

24.   A method of diagnosing disease or predisposition to disease in an individual not showing significant signs of disease, comprising use of the database of Claim 1

5

ADD A1